STATE OF WASHINGTON
DEPARTMENT OF HEALTH
Olympia, Washington 98504

June 24, 1999

Dr. Scott Davis, Director
Hanford Thyroid Disease Study
Fred Hutchinson Cancer Research Center
1124 Columbia Street,
MP-425
Seattle, Washington 98104

SUBJECT: DRAFT HANFORD THYROID DISEASE STUDY (HTDS) REPORT

Dear Dr. Davis:

The Washington State Department of Health congratulates you and your staff for your achievement in designing and implementing the Hanford Thyroid Disease Study (HTDS). It is a study of extreme complexity. You were successful in contacting people selected for the sample and the participation rate among those contacted was extremely high, given the amount of time that had elapsed between the potential exposure and the study. Your preliminary activities to test critical assumptions and field strategies were invaluable in contributing to a well-designed and well-conducted study in spite of several inherent limitations due to the need to reconstruct exposures from many years ago. In short, we believe you did an excellent job given the parameters within which you had to work.

We are confident that if there were adverse thyroid health effects related to exposure to iodine 131 from past releases at the Hanford Nuclear Site, the excess risk of thyroid disease to the exposed population was not large. However, we believe that the study may not have been able to detect relatively small increases in thyroid disease related to exposure to iodine 131. In theory, you seem to have the power "to detect reasonably small effects..." as stated in the last paragraph of the Discussion section. However, your ability to discern small effects may be diminished by both random misclassification related to exposure and your choice of a comparison group.

In fact, we have difficulty in knowing what you mean by relatively small. On page 3 of the Introduction, you state that you had sufficient power to detect an increase of 5% in thyroid neoplasia per Gray. We do not know whether a 5% increase means from 1.3% (i.e., the rate in the low exposure group, page 78, Section VIII) to 6.3% or from 1.3% to 1.37%. A clarification of this point would help us better understand what you mean by "relatively small."

MISCLASSIFICATION OF EXPOSURE

As you describe on page 9 under C.3 of the Discussion section, random misclassification related to exposure will tend to produce findings of no difference between relatively highly exposed individuals and participants with relatively lower exposures. As you explain in section C.3, there are two major sources of error that may lead to misclassification: error in dietary recall and error in the Hanford Dose Reconstruction (HEDR) estimates.

Misclassification related to dietary recall

- The report of the HTDS does not include a section on dietary intake methodology, including the strengths and weaknesses of different approaches to obtaining dietary intake information, such as use of proxy respondents and accuracy of reporting intake from many years ago. While we know of no studies addressing the reliability of dietary recall that are exactly comparable to the methods used in the HTDS, we know of several studies that may be relevant.

  For example, Dwyer et al. noted relatively good recall (correlation coefficient of 0.5) of dietary intake of dairy products among people in their 50s when asked to report what they ate when they were 30 years old. However, correlations between the original reports and recall were much smaller (correlation coefficients between 0.2 and 0.3) when the same people were asked to report what they ate as adolescents and children. (Dwyer JT, Gardner J, Halvorsen K, Krall EA, Cohen A and Valadian I. Memory of food intake in the distant past. American Journal of Epidemiology. 130(5): 1033 -1046, 1989.)

  Fraser et al. noted a correlation coefficient of 0.4 for dietary recall of whole milk by people in their 70s recalling what they had reported 20 years earlier. (Fraser GE, Linsted KD, Knutsen SY, Beeson WL, Bennet H and Shavlik DJ. Validity of dietary recall over 20 years among California Seventh-day Adventists. American Journal of Epidemiology, 148: 810-818, 1998.)

  While the authors of the studies cited above believe the correlation coefficients of 0.4-0.5 found in their studies indicate the legitimacy of asking about dietary habits in the distant past, we are concerned that the degree of misclassification may be greater

than that represented by these correlation coefficients. The correlations of 0.4-0.5 are for older adults reporting on their habits as younger adults. It may be that milk drinking among the ages represented in the above studies (i.e., between 30 to 50 years old and 50 to 70 years old) is more consistent and thus easier to report than in the HTDS, where people were asked to report on their own or their children's milk intake from 40 to 50 years ago.

Discussion and citation of this literature in the Background or Discussion sections may help the reader gain perspective on the likelihood, direction and magnitude of misclassification due to errors in reporting of dietary intake. For example, based on Armstrong et al., if the true odds of disease are twice as high among those in the highest quartile of exposure compared to those in the lowest quartile of exposure (i.e., an odds ratio of 2) and the correlation between the true dietary intake and reported dietary intake is 0.5, the observed odds ratio would be 1.5. (Armstrong BK, White E. and Saracci R. *Principles of Exposure Measurement in Epidemiology.* Oxford University Press, 1994, pp 59-63.) As the correlation between the true and reported dietary intake decreases, the discrepancy between the true and observed risk of disease will increase.

- It is difficult to get an overview of who provided dietary information for what proportion of the final HTDS study population. We urge you to summarize this information in Section V or Section VIII. The literature indicates that self-report of adults on dietary habits from childhood is not reliable (Dwyer op. cit.) While reports from mothers may be more reliable, we question the reliability of reports from siblings and other relatives. (Since we have not identified literature addressing these issues, our concern is not grounded in scientific literature, but rather represents our best guess.)

  From pages 83, 85 and 117 of Section V, it appears that for almost 40% of the study population you used self-reported milk intake; for another 45% of the population, you used mother's report; and for 15% of the sample, you used other reporters, such as sisters and other relatives. (We get slightly different figures on the distribution of CATI and expanded IPI from page 15 of Section VIII.)

- An explanation of the sensitivity of the model to different levels of reporting of dietary intake might also help the reader in assessing the importance of misclassification due to inaccuracies in reporting of diet. For example, what is the calculated dose for reporting 3 glasses of milk per day compared to 4 glasses or 6 glasses?

- While you assessed thyroid disease using exposure based on both the reported dietary intake and the HEDR reference diet, you did not seem to include any comparisons of the reported diet to the reference diet. Again, this information might help the reader

assess the importance of misclassification of exposure due to inaccuracy in reporting of dietary intake.

- We applaud the research team for consulting with Dr. Dillman to try to improve recall of events from 40 to 45 years ago. However, we know of none of Dr. Dillman's published research that addresses the reliability and validity of dietary recall over lengthy time periods. You have not included citations for Dr. Dillman and so the reader has no method of assessing the claim, "The cognitive approach to interviewing undoubtedly helped in improving the accuracy of recall..." (p. 10 Section IX).

- The study methods did not seem to include any attempt to ask questions that would allow for an assessment of internal reliability of dietary recall. For example, at the end of the interview, the participants could have been asked again about the amount of milk, eggs and butter the subject consumed at several different time periods. One would expect consistency over the relatively short period of time during which the interview occurred. Lack of consistency might speak to serious misclassification errors. Alternatively, current total caloric intake could have been assessed using a food frequency questionnaire. This approach may have allowed for detection of gross over- or underreporting for which default measures might have been more accurate than reported intake.

- Because of concern about misclassification of exposure due to inaccuracies in dietary recall, the investigators used several different models to assess consumption, as explained on page 10 of Section IX. In one model, default values for a small number of participants (62 per our calculations) for whom individual dietary intake was not available were changed. In the other model, all participants were assigned default values based on residence. These models may very well substitute one data set with random misclassification with another set of data with an equal amount of misclassification. Therefore, we cannot agree with the conclusion that "this provides some assurance that the absence of a dose-response found in the primary analyses is not due to misclassification of exposure introduced by difficulties in recall from the distant past." (p. 10, Section IX) In fact, if the investigators have confidence in the assignment of dose based on default values, they could have simplified the study by utilizing those values to begin with.

<u>Misclassification related to calculations using the HEDR model</u>

- It is not clear whether the soil deposition estimates were based on meteorological data from 1944-1947 or from the 1980s (page 4, Section IV, Study Design). If the meteorological data are from the 1980s, it is important to know whether the releases from Hanford were relatively consistent over time or whether they occurred sporadically. For example, with reference to Figure IV-3, were the releases evenly distributed over the whole month or did they occur on one or two days of the month.

If the meteorological data are from the 1980s, we are more comfortable with using these data if the pattern is the former.

- The HEDR project included both airborne releases of iodine 131 and exposures related to the Columbia River. The HTDS only included exposures to airborne iodine 131. Therefore, for heavy Columbia River users, the estimated dose may have been lower than the actual dose. While this may not affect the classification of a large portion of the cohort, it is another source of misclassification adding to those sources already identified. A discussion of this exposure pathway might help the reader assess the importance of omitting this potential exposure from consideration in calculating doses. Additionally, given that you did not include potential exposures associated with the Columbia River, we think that you need to be clearer that you are assessing only exposures related to airborne releases of iodine 131. Perhaps wording in the title of the study to this effect would be helpful in clarifying this point.

- Our understanding from the HTDS manuscript is that it is not possible to validate the estimates of soil deposition of I-131 used by the HEDR model. Our experience with modeling is that external validation is important in assuring the accuracy of the model. Our concern about the possibility of inaccuracies in the information used by HEDR is reinforced by recent information requiring a recalculation of dose based on the new information. If there have been similar models that have been validated by external measurements, it might be helpful to include that information so that the reader can assess the potential for errors. It might also be informative to include a discussion of some of the international model intercomparisons that have included the HEDR model (IAEA 1999, "Model testing using data on iodine 131 released from Hanford", BIOMASS Theme 2 Dose Reconstruction Working Group, International Atomic Energy Agency, Vienna, Austria. IN PRESS).

- As discussed in the previous on dietary recall, alternative representations of exposure not utilizing the HEDR model will also contain random misclassification and so similar findings using alternative models do not assure us that the findings are robust.

CHOICE OF COMPARISON GROUP

The decision to use a low-dose rather than a no-dose comparison group may also limit the study's ability to detect small effects due to exposure to iodine 131. By analogy, it is more difficult to detect effects of smoking when looking at light smokers compared to heavy smokers than when comparing smokers to non-smokers. A more detailed explanation of the reasons to pursue the low dose/high dose comparison would be helpful.

NEVADA TEST SITE FINDINGS

A discussion of the results of studies from the Nevada Test Site (NTS) may help the reader place the findings from the HTDS into perspective. On page 8 of Section II, you provide information that indicates associations between thyroid disease and exposures from the NTS. From the information in section B.2 of Section VIII, it appears that many participants had doses at the level of participants in the NTS study. An explanation of the reasons for differences in the findings of the NTS and the HTDS would be helpful in assessing the HTDS results.

CONCLUSIONS

The Department commends the researchers on their well-designed and well-conducted study in spite of inherent limitations due to the need to reconstruct doses from exposures occurring many years ago. While we understand the complexities of the study, we do have concerns about the ability to accurately estimate radiation doses. We are concerned that these inaccuracies may have resulted in random misclassification of exposure, resulting in an inability of the study to detect small increases in risk of thyroid disease. The use of a low-dose comparison group, instead of a no-dose comparison group, may contribute to making detection of small increases in thyroid disease problematic. Even with these concerns, the Department is confident that the study provides evidence that if releases from Hanford increased thyroid disease, the increase was relatively small on a population basis.

Sincerely,

Juliet VanEenwyk
Acting State Epidemiologist

John Erickson,Director
Division of Radiation Protection

Cc: Dr. Maxine Hayes, Acting State Health Officer, Washington State Department of Health
Dr. Paul Garbe, CDC
Dr. Steven L. Simon, NAS
Ms. Laura Chenet Leonard, Oregon Health Division
Ms. Elke Shaw-Tulloch, Idaho Division of Health